

Generic versus Salient Region-based Partitioning for Local Appearance Face Recognition

Hazım Kemal Ekenel and Rainer Stiefelhagen

Computer Science Department, Universität Karlsruhe (TH)
Am Fasanengarten 5, Karlsruhe 76131, Germany
{[ekenel](mailto:ekenel@ira.uka.de),[stiefel](mailto:stiefel@ira.uka.de)}@ira.uka.de
<http://isl.ira.uka.de/cvhci>

Abstract. In this paper, we investigate different partitioning schemes for local appearance-based face recognition. Five different salient region-based partitioning approaches are analyzed and they are compared to a generic partitioning scheme. Extensive experiments have been conducted on the AR, CMU PIE, FRGC, Yale B, and Extend Yale B face databases. The experimental results show that generic partitioning provides better performance than salient region-based partitioning schemes.

1 Introduction

Since the beginning of 1990s, appearance-based holistic approaches have been dominating the face recognition research [1–3]. Although local appearance information, in other words, using salient regions, has been shown to be superior to the holistic information in [4, 5], interestingly face recognition research has focused on holistic approaches and local appearance based face recognition has been ignored in a great extent. It has not had as much impact as the holistic approaches, and compared to the plethora of the holistic methods, only a few techniques have been proposed to perform local appearance-based face recognition. The main reason for this is that the initial local appearance based approaches [4, 5] require the detection of salient features i.e. eyes- which may not be an easy task. Moreover, erroneous detection of these local regions leads to severe performance drops.

Recently, local appearance-based face recognition approaches have been attracting growing interest [6–14]. In [6, 14], the local facial regions are located by a support vector machine (SVM) and the combined local features are classified again with SVM. In [7], the face image is divided into rectangular sub-images without considering any salient regions, and the eigenfaces approach is then performed on each of these sub-images. In [8], the features extracted from local parts are modelled using GMM which are estimated by benefiting relevance adaptation (RA). In [11], the face image is partitioned into several local regions and each local region is represented by linear discriminant analysis (LDA). To combine the features extracted from each local region, another LDA is used. A combination of DCT and LDA is applied on facial components in [9]. In [13], local

binary pattern based face recognition is proposed. In this approach the image is divided into rectangular blocks and each block is represented by histogram of the local binary patterns (LBP). In [10, 12], a generic face representation approach is introduced for local appearance-based face recognition. Local features, that are derived from DCT coefficients, are used for representing the local regions. The obtained local features are combined by conserving the spatial information. The proposed approach has been also utilized for 3-D and video-based face recognition [15, 16].

In this study, we investigate the effect of different salient region-based partitioning schemes to the performance of the local appearance based face recognition (LAFR) approach. Normally, the proposed LAFR technique partitions an aligned face image into 8×8 pixels resolution non-overlapping blocks without considering any salient regions [10, 12]. In addition to this generic partitioning, we test five different salient region-based partitioning schemes. Experimental results on AR [17], CMU PIE [18], FRGC [19], Yale B [20] and Extended Yale B [21] face databases show that generic partitioning provides better performance than salient region-based partitioning schemes.

The organization of the paper is as follows. In Section 2, local appearance-based face recognition is explained. Utilized partitioning approaches are introduced in Section 3. Experimental results are presented and discussed in Section 4. Finally, in Section 5, conclusions are given.

2 Local appearance-based face recognition

Local appearance-based face recognition is based on statistical representations of the non-overlapping local facial regions and their combination at the feature level. The underlying idea is to utilize local information while preserving the spatial relationships. In [10], the discrete cosine transform (DCT) is proposed to be used to represent the local regions. Its compact representation ability is superior to that of the other widely used input independent transforms like Walsh-Hadamard transforms. Although Karhunen-Loève transform (KLT) is known to be the optimal transform in terms of information packing, its data dependent nature makes it infeasible to use it for some practical tasks. Furthermore, DCT closely approximates the compact representation ability of the KLT, which makes it very useful for representation both in terms of information packing and in terms of computational complexity. In addition, DCT has been shown to be a better representation method for modeling the local facial appearance compared to KLT in terms of face recognition performance [10].

Local appearance-based face recognition approach is proposed as a generic, practical and robust face recognition algorithm. Feature extraction using local appearance-based face representation can be summarized as follows: A detected and normalized face image is divided into blocks of 8×8 pixels resolution. The reason for having 8×8 pixels block size is to provide sufficient compactness on one hand, and to keep stationarity within the block on the other hand. The DCT is applied on each block. Then, the obtained DCT coefficients are ordered using

zig-zag scan pattern. From the ordered coefficients, M of them are selected and normalized according to a feature selection and feature normalization strategy resulting in an M -dimensional local feature vector [12]. Finally, the DCT coefficients extracted from each block are concatenated to construct the global feature vector. For details of the algorithm please see [10, 12].

3 Generic vs. Salient Region-based Partitioning

In this study, five different salient region-based partitioning schemes, that are derived from the previous modular/component/patch based studies [4, 5, 9, 11, 14], are compared to the generic partitioning of the local appearance-based face recognition approach. In the implementation, the salient regions are divided into 8×8 pixels resolution non-overlapping blocks and the DCT is applied on each block. From the DCT coefficients that are ordered according to the zig-zag scan pattern, ten of them are selected by omitting the first DCT coefficient and selecting the following first ten of them. The selected coefficients are divided to their standard deviations. Afterwards, the local feature vector is normalized to the unit norm. The overall feature vector for a salient region is constructed by concatenating the local feature vectors that are extracted from the blocks of the corresponding salient region. The feature vector of the combined regions is generated by concatenating the local feature vectors of each region.

The generic partitioning is performed by dividing the face image into 8×8 pixel resolution non-overlapping blocks as depicted in Figure 1. The first salient region-based partitioning scheme (P1) is similar to the one in [4]. It consists of three regions: Eyes, nose and mouth. A sample image, illustrating this partitioning scheme, is given in Figure 2. The second partitioning scheme (P2) is from [5]. Four salient regions are used for face recognition: Left eye, right eye, nose and mouth. The partitioning on a sample image is shown in Figure 4. The third partitioning scheme (P3) is derived from [9]. Larger left eye and right eye regions that contain partially the nose and some parts below the eyes, and the nose region are the salient regions used in this partitioning scheme. The partitioning on a sample image is illustrated in Figure 6. The fourth partitioning scheme is an approximation of the one in [11]. It has five regions: forehead, left eye, right eye, lower left and right parts of the face. The partitioning on a sample image can be seen in Figure 8. The fifth partitioning scheme is derived from [14]. There are 14 learned components as shown in Figure 10.

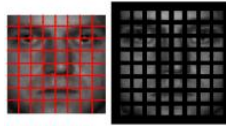


Fig. 1. Generic partitioning scheme.

4 Experiments

The performance of the algorithm is tested on five different benchmarks, that are widely used in the face recognition research community. The face recognition grand challenge (FRGC) version 2 data set [19] is used to test the algorithm under controlled and uncontrolled conditions. For controlled conditions, 120 individuals from the FRGC experiment 1 data set are selected, who have at least ten images both in fall 2003 and spring 2004 recordings. The images from fall 2003 recordings are used for training and the ones from spring 2004 are used for testing. For uncontrolled conditions, the same experimentation setup is used, but this time the images are selected from the FRGC experiment 4 data set. The AR face database [17] is utilized for testing against partial face occlusion. 550 face images of 110 subjects are selected from this database. Separate experiments were conducted for upper and lower face occlusion. For both cases, a single face image that does not contain any occlusion is used for training. To assess the performance against upper face occlusion, a single face image that contains sun glasses is used. For lower face occlusion experiments, a single face image that contains a scarf is used. In order to test the algorithm against illumination variations, experiments are conducted on the CMU PIE [18] and the Yale/Extended Yale B face databases [20, 21]. There are 68 subjects in the CMU PIE database. 21 images per subject are chosen from the illumination set for the experiments. From these images, only the frontally illuminated face image is used for training. The remaining twenty images are used for testing. The Yale face database B contains ten subjects under 64 different illumination conditions. The first subset that has close to frontal illumination is used for training. For testing, subsets 2, 3, 4 and 5 are used. These subsets contain 12, 12, 14, 19 images per person, respectively. With increasing subset number, the illumination variations become stronger. The extended Yale face database B contains Yale face database B and has the same organization. It has 28 additional subjects making 38 subjects in total. Overall, there are 15 training-testing combinations. Two from the FRGC ver. 2 database, one for controlled and one for uncontrolled case; one from the CMU PIE; four from the AR, two for lower face occlusion and two for upper face occlusion -one of them is within session and the other one is between session experiment. Finally, eight from the Yale B and Extended Yale B face databases. Four from each one.

The obtained results using P1 partitioning scheme with the individual components and the combined representation on each experimental setup are shown in Figure 3. The correspondences between the abbreviations on the x-axis and the experiments are as follows: F1: FRGC under controlled conditions, F4: FRGC under uncontrolled conditions, CP: CMU PIE, A1sc: AR, occlusion with scarf, within session, A1sun: AR, occlusion with sunglasses, within session, Aisc: AR, occlusion with scarf, between session, Aisun: AR, occlusion with sunglasses, between session, Y2,Y3,Y4,Y5: Yale subset 2,3,4,5 and EY2,EY3,EY4,EY5: Extended Yale subset 2,3,4,5, respectively. The best results are obtained with the combined representation except on the Aisun experiment, where mouth region provides the highest correct classification rate. Eye region is found to be the

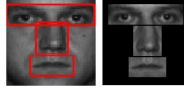


Fig. 2. Salient regions obtained with the P1 partitioning scheme.

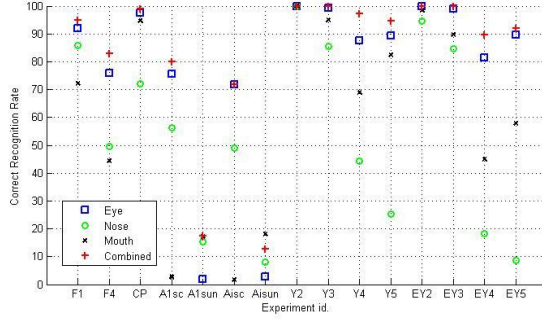


Fig. 3. Correct identification rates obtained with the P1 partitioning scheme.

second best performing region, except in the experiments where upper face occlusion exists. On Y2 and EY2, 100% correct recognition rate is achieved just using the eye region. Depending on the experimental setup, either nose region or mouth region comes the third. In the experiments with high illumination variations, such as CP, Y4, Y5, EY4 and EY5, mouth region is found to be more useful for identification than the nose region. This is expected, since in the case of illumination variation due to cast shadows the appearance of nose region is affected severely. The other reason for this outcome is the lack of expression variations in the used data sets for these experiments. Mouth region works also better in the experiments that contain upper face occlusion, namely, A1sun and Aisun experiments. Having sunglasses decreases the amount of the discriminative information that the nose region contains. In F1, F4, A1sc and Aisc experiments nose region reaches higher recognition rates than the mouth region. The expression variations in F1 and F4 experiments deteriorate the performance of the mouth region. Obviously, in the case of lower face occlusion, the mouth region has no use.

The results obtained by using P2 partitioning scheme can be seen from Figure 5. The outcomes are similar to the ones obtained with the P1 partitioning scheme. Combined representation achieves the best results. Eye regions have the second place. There is no big difference in left and right eye regions' correct classification rates. The same observations are valid for the nose and mouth regions. The only difference is observed on the CP experiment where the mouth region performs slightly better than the eye regions. It can be also observed that the performance difference between the mouth region and the eye regions is less than the one attained on the experiments that contain illumination variations with P1 partitioning scheme. The reason is, the region that contains both of the eyes have more discriminative power than the individual eye regions.

The correct identification rates that are achieved with the P3 partitioning scheme are presented in Figure 7. Combined representation attains the highest

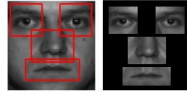


Fig. 4. Salient regions obtained with the P2 partitioning scheme.

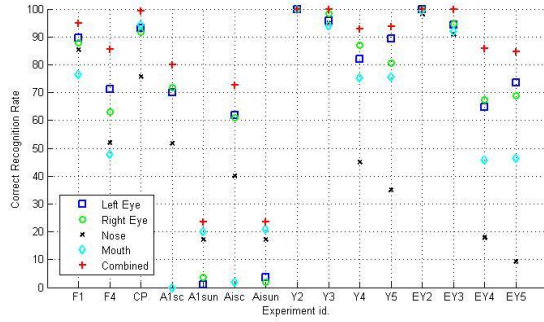


Fig. 5. Correct identification rates obtained with the P2 partitioning scheme.

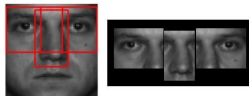


Fig. 6. Salient regions obtained with the P3 partitioning scheme.

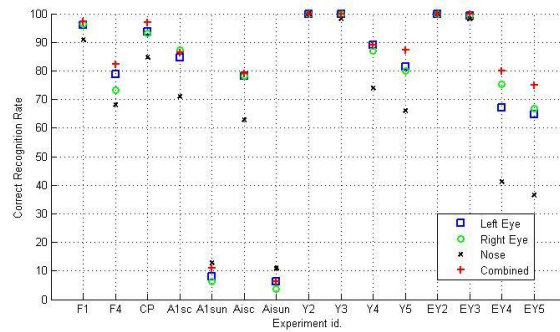


Fig. 7. Correct identification rates obtained with the P3 partitioning scheme.

correct recognition rates in most of the experiments. On A1sun and Aisun nose region achieves the best results, whereas on A1sc right eye region outperforms the others. Once more time, it has been observed that, except upper face occlusion, eye regions contain more discriminative power than the nose region.

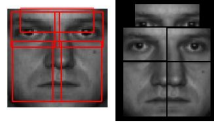


Fig. 8. Salient regions obtained with the P4 partitioning scheme.

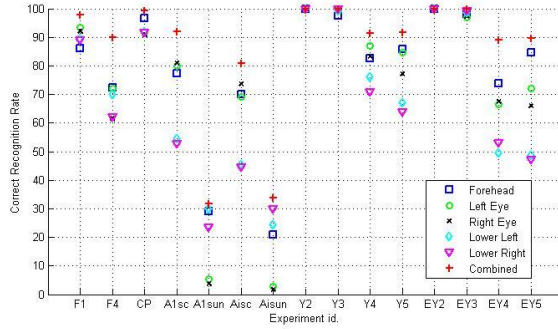


Fig. 9. Correct identification rates obtained with the P4 partitioning scheme.

Figure 9 shows the correct identification rates obtained by using the P4 partitioning scheme. The best performance is always achieved with the combined representation. On the experiments with lower face occlusion, as expected, lower face regions perform poorly and on the experiments with upper face occlusion, eye regions perform poorly. In most of the cases forehead region achieves higher correct recognition rates compared to the other salient regions on the experiments that contain large illumination variations, since this region is less affected from the changes in lighting. Both the eye regions and lower facial parts contain partially the nose region which makes them sensitive to the changes in appearance due to cast shadows.

Correct identification rates attained by using the P5 partitioning scheme are given in Figure 11. Most of the time the best performance is achieved with the combined representation. Only on A1sc, Aisc experiments nose bridge region, and on A1sun right cheek region performs better. Depending on the experimental condition, performance order of the facial parts changes. Besides the experiments with upper face occlusion, right eye, left eye, right eyebrow and left eyebrow regions consistently achieve high classification rates.

The comparison of combined representation of different partitioning schemes is shown in Figure 12. Generic partitioning is found to be superior to the salient region-based partitioning in most of the cases. For example, on F4, the performance is 83.0% with P1, 85.6% with P2, 82.3% with P3, 89.8% with P4, 82.8% with P5 and 90.8% with generic partitioning. Only on the Y4 experiment P1 partitioning scheme outperforms generic partitioning. However, on EY4, which

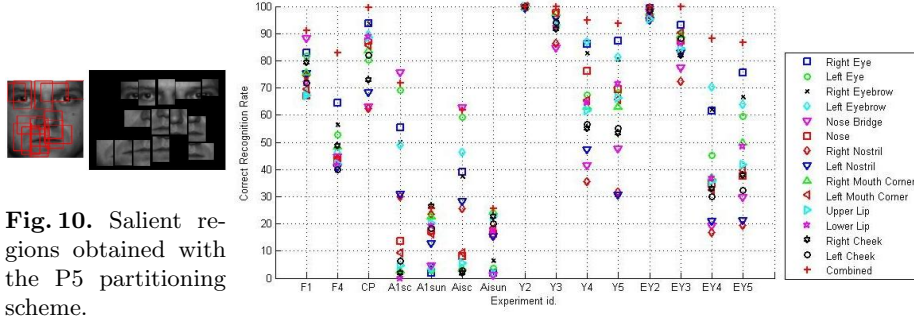


Fig. 10. Salient regions obtained with the P5 partitioning scheme.

Fig. 11. Correct identification rates obtained with the P5 partitioning scheme.

contains Y4 as a subset, generic partitioning provides better results. P4 partitioning scheme also provides consistently high results. On the experiments that contain large illumination variations, P3 partitioning scheme is found to be the poorest performing one. The reason is on each part nose is included to some extend, which makes it sensitive to the cast shadows. P1 performs better than P2 on these experiments again due to its less sensitivity to the cast shadows.

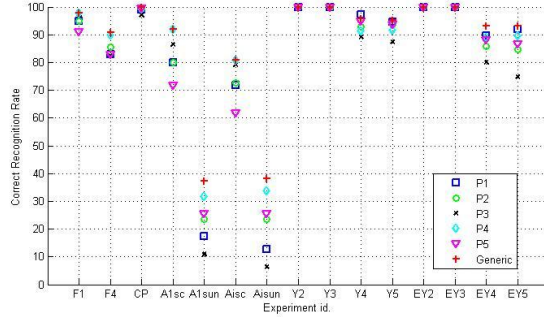


Fig. 12. Correct identification rates obtained with the combined representations.

5 Conclusions

In this study, we evaluate different salient region-based partitioning schemes for local appearance-based face recognition algorithm, which is originally proposed as a generic approach and partitions an aligned face image into 8×8

pixels resolution non-overlapping blocks without considering any salient regions. We compare five different salient region-based partitioning schemes. It has been found that generic partitioning provides higher correct recognition rates than salient region-based partitioning approaches. Thus, justifying that there is no need to focus on salient regions and perform salient region-based partitioning.

Acknowledgement

This study is partially funded by the German Research Foundation (DFG) under Sonderforschungsbereich SFB 588 —Humanoid Robots— and by OSEO, French State agency for innovation, as part of the Quaero Programme. The authors would like to thank Dr. Bernd Heisele and Dr. Tae-Kyun Kim for providing information about the partitioning schemes that they have used.

References

1. Turk, M., Pentland, A.: Eigenfaces for recognition. *Journal of Cognitive Neuroscience* **3**(1) (1991) 71–86
2. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Trans. on PAMI* **19**(7) (1997) 711–720
3. Moghaddam, B., Jebara, T., Pentland, A.: Bayesian face recognition. *Pattern Recognition* **33**(11) (2000) 1771–1782
4. Brunelli, R., Poggio, T.: Face recognition: Features versus templates. *IEEE Trans. on PAMI* **15**(10) (1993) 1042–1052
5. Pentland, A., Moghaddam, B., Starner, T.: View-based and modular eigenspaces for face recognition. In: *Proc. of IEEE CVPR*. (1994) 84–91
6. Heisele, B., Ho, P., Wu, J., Poggio, T.: Face recognition: component-based versus global approaches. *Computer Vision and Image Understanding* **91**(1-2) (2003) 6–21
7. Gottumukkal, R., Asari, V.K.: An improved face recognition technique based on modular PCA approach. *Pattern Recognition Letters* **25**(4) (2004) 429–436
8. Lucey, S., Chen, T.: A GMM parts based face representation for improved verification through relevance adaptation. In: *Proc. of IEEE CVPR*. (2004) 855–861
9. Lee, H.J., Kim, H.J., Kim, W.Y.: Face recognition using component-based DCT/LDA. In: *Proc. of IWAIT*. (2005)
10. Ekenel, H.K., Stiefelhagen, R.: Local appearance based face recognition using discrete cosine transform. In: *Proc. of EUSIPCO*. (2005)
11. Kim, T.K., Kim, H., Hwang, W., Kittler, J.: Component-based LDA face description for image retrieval and MPEG-7 standardisation. *Image and Vision Computing* **23**(7) (2005) 631–642
12. Ekenel, H.K., Stiefelhagen, R.: Analysis of local appearance-based face recognition: Effects of feature selection and feature normalization. In: *Proc. of IEEE CVPR Biometrics Workshop*. (2006)
13. Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: application to face recognition. *IEEE Trans. on PAMI* **28**(12) (2006) 2037–2041
14. Heisele, B., Serre, T., Poggio, T.: A component-based framework for face detection and identification. *International Journal of Computer Vision* **74**(2) (2007) 167–181

15. Ekenel, H.K., Gao, H., Stiefelhagen, R.: 3-D face recognition using local appearance-based models. *IEEE Trans. on Information Forensics and Security* **2**(3) (2007) 630–635
16. Stallkamp, J., Ekenel, H.K., Stiefelhagen, R.: Video-based face recognition on real-world data. In: *Proc. of IEEE ICCV*. (2007) 1–8
17. Martinez, A.M., Benavente, R.: The AR face database (1998) CVC Tech. Report #24.
18. Sim, T., Baker, S., Bsat, M.: The CMU pose, illumination, and expression (PIE) database. *IEEE Trans. on PAMI* **25**(12) (2003) 1615–1618
19. Phillips, P.J., Flynn, P.J., Scruggs, T., Bowyer, K.W., Chang, J., Hoffman, K., Marques, J., Min, J., Worek, W.: Overview of the face recognition grand challenge. In: *Proc. of IEEE CVPR*. (2005) 947–954
20. Georgiades, A.S., Belhumeur, P.N., Kriegman, D.J.: From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. on PAMI* **23**(6) (2001) 643–660
21. Lee, K.C., Ho, J., Kriegman, D.J.: Acquiring linear subspaces for face recognition under variable lighting. *IEEE Trans. on PAMI* **27**(5) (2005) 684–698